

1012

#2

OIPE

RAW SEQUENCE LISTING

DATE: 09/25/2001

PATENT APPLICATION: US/09/863,475

TIME: 10:33:32

Input Set : N:\Crf3\RULE60\09863475.txt

Output Set: N:\CRF3\09252001\I863475.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:

6 (i) APPLICANT: LOWE, JOHN B.

8 (ii) TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
 9 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
 10 GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
 11 OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

13 (iii) NUMBER OF SEQUENCES: 14

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 17 P.C.

18 (B) STREET: 1755 Jefferson Davis Highway, Fourth Floor

19 (C) CITY: Arlington

20 (D) STATE: Virginia

21 (E) COUNTRY: U.S.A.

22 (F) ZIP: 22202

ENTERED

24 (v) COMPUTER READABLE FORM:

25 (A) MEDIUM TYPE: Floppy disk

26 (B) COMPUTER: IBM PC compatible

27 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

28 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

30 (vi) CURRENT APPLICATION DATA:

C--> 31 (A) APPLICATION NUMBER: US/09/863,475

C--> 32 (B) FILING DATE: 24-May-2001

33 (C) CLASSIFICATION:

35 (vii) PRIOR APPLICATION DATA:

36 (A) APPLICATION NUMBER: 07/914,281

37 (B) FILING DATE:

41 (viii) ATTORNEY/AGENT INFORMATION:

42 (A) NAME: Lavalleye, Jean-Paul M. P.

43 (B) REGISTRATION NUMBER: 31,451

44 (C) REFERENCE/DOCKET NUMBER: 2363-060-55

46 (ix) TELECOMMUNICATION INFORMATION:

47 (A) TELEPHONE: (703)521-4500

48 (B) TELEFAX: (703)486-2347

49 (C) TELEX: 248855 OPAT UR

52 (2) INFORMATION FOR SEQ ID NO: 1:

54 (i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 2043 base pairs

56 (B) TYPE: nucleic acid

57 (C) STRANDEDNESS: unknown

58 (D) TOPOLOGY: unknown

60 (ii) MOLECULE TYPE: cDNA

62 (iv) ANTI-SENSE: NO

66 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

68 AGGAAACCTG CCATGGCCTC CTGGTGAGCT GTCCTCATCC ACTGCTCGCT GCCTCTCCAG 60

70 ATACTCTGAC CCATGGATCC CCTGGGTGCA GCCAAGCCAC AATGGCCATG GCGCCGCTGT 120

RAW SEQUENCE LISTING

DATE: 09/25/2001

PATENT APPLICATION: US/09/863,475

TIME: 10:33:32

Input Set : N:\Crf3\RULE60\09863475.txt

Output Set: N:\CRF3\09252001\I863475.raw

```

72 CTGGCCGCAC TGCTATTTCA GCTGCTGGTG GCTGTGTGTT TCTTCTCCTA CCTGCGTGTG 180
74 TCCCAGACAG ATGCCACTGG ATCCCCTAGG GCTCCCAGTG GGTCTCTCCG ACAGGACACC 240
76 ACTCCCACCC GCCCCACCTT CCTGATCCTG CTATGGACAT GGCCTTTCCA CATCCCTGTG 300
78 GCTCTGTCCC GCTGTTTACA GATGGTGCCC GGCACAGCCG ACTGCCACAT CACTGCCGAC 360
80 CGCAAGGTGT ACCCACAGGC AGACACGGTC ATCGTGCACC ACTGGGATAT CATGTCCAAC 420
82 CCTAAGTCAC GCCTCCCACC TTCCCCGAGG CCGCAGGGGC AGCGCTGGAT CTGGTTCAAC 480
84 TTGGAGCCAC CCCCTAACTG CCAGCACCTG GAAGCCCTGG ACAGATACTT CAATCTCACC 540
86 ATGTCCTACC GCAGCGACTC CGACATCTTC ACGCCCTACG GCTGGCTGGA GCCGTGGTCC 600
88 GGCCAGCCTG CCCACCCACC GCTCAACCTC TCGGCCAAGA CCGAGCTGGT GGCCTGGGCG 660
90 GTGTCCAAC TGAAGCCGGA CTCAGCCAGG GTGCGCTACT ACCAGAGCCT GCAGGCTCAT 720
92 CTCAAGGTGG ACGTGTACGG ACGCTCCCAC AAGCCCCTGC CCAAGGGGAC CATGATGGAG 780
94 ACGCTGTCCC GGTACAAGTT CTACCTGGCC TTCGAGAACT CCTTGCACCC CGACTACATC 840
96 ACCGAGAAGC TGTGGAGGAA CGCCCTGGAG GCCTGGGCGG TGCCCGTGGT GCTGGGCCCC 900
98 AGCAGAAGCA ACTACGAGAG GTTCTTGCCA CCCGACGCCT TCATCCACGT GGACGACTTC 960
100 CAGAGCCCCA AGGACCTGGC CCGGTACCTG CAGGAGCTGG ACAAGGACCA CGCCCGCTAC 1020
102 CTGAGCTACT TTCGCTGGCG GGAGACGCTG CGGCCTCGCT CCTTCAGCTG GGCCTGGAT 1080
104 TTCTGCAAGG CCTGCTGGAA ACTGCAGCAG GAATCCAGGT ACCAGACGGT GCGCAGCATA 1140
106 GCGGCTTGGT TCACCTGAGA GGCCGGCATG GTGCCTGGGC TGCCGGGAAC CTCATCTGCC 1200
108 TGGGGCCTCA CCTGCTGGAG TCCTTTGTGG CCAACCTCT CTCTTACCTG GGACCTCACA 1260
110 CGCTGGGCTT CACGGCTGCC AGGAGCCTCT CCCCTCCAGA AGACTTGCCT GCTAGGGACC 1320
112 TCGCCTGCTG GGGACCTCGC CTGTTGGGGA CCTCACCTGC TGGGGACCTC ACCTGCTGGG 1380
114 GACCTTGGCT GCTGGAGGCT GCACCTACTG AGGATGTGCG CGGTCGGGGA CTTTACCTGC 1440
116 TGGGACCTGC TCCAGAGAC CTTGCCACAC TGAATCTCAC CTGCTGGGGA CCTCACCTG 1500
118 GAGGGCCCTG GGCCCTGGGG AACTGGCTTA CTTGGGGGCC CACCCGGGAG TGATGGTTCT 1560
120 GGCTGATTTG TTTGTGATGT TGTTAGCCGC CTGTGAGGGG TGCAGAGAGA TCATCACGGC 1620
122 ACGGTTTCCA GATGTAATAC TGCAAGGAAA AATGATGACG TGTCTCCTCA CTCTAGAGGG 1680
124 GTTGGTCCCA TGGGTAAAGA GCTCACCCCA GGTTCCTACC TCAGGGGTTA AGAGCTCAGA 1740
126 GTTCAGACAG GTCCAAGTTC AAGCCCAGGA CCACCACTTA TAGGGTACAG GTGGGATCGA 1800
128 CTGTAAATGA GGACTTCTGG AACATTCCAA ATATTCTGGG GTTGAGGGAA ATTGCTGCTG 1860
130 TCTACAAAAT GCCAAGGGTG GACAGGCGCT GTGGCTCACG CCTGTAATTC CAGCACTTTG 1920
132 GGAGGCTGAG GTAGGAGGAT TGATTGAGGC CAAGAGTTAA AGACCAGCCT GGTCAATATA 1980
134 GCAAGACCAC GTCTCTAAAT AAAAAATAAT AGGCCGGCCA GGAAAAAAAA AAAAAAAAAA 2040
136 AAA 2043

```

138 (2) INFORMATION FOR SEQ ID NO: 2:

140 (i) SEQUENCE CHARACTERISTICS:

141 (A) LENGTH: 361 amino acids

142 (B) TYPE: amino acid

143 (D) TOPOLOGY: unknown

145 (ii) MOLECULE TYPE: protein

149 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

151 Met Asp Pro Leu Gly Ala Ala Lys Pro Gln Trp Pro Trp Arg Arg Cys
152 1 5 10 15
154 Leu Ala Ala Leu Leu Phe Gln Leu Leu Val Ala Val Cys Phe Phe Ser
155 20 25 30
157 Tyr Leu Arg Val Ser Arg Asp Asp Ala Thr Gly Ser Pro Arg Ala Pro
158 35 40 45
160 Ser Gly Ser Ser Arg Gln Asp Thr Thr Pro Thr Arg Pro Thr Leu Leu
161 50 55 60
163 Ile Leu Leu Trp Thr Trp Pro Phe His Ile Pro Val Ala Leu Ser Arg

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/863,475

DATE: 09/25/2001
TIME: 10:33:32

Input Set : N:\Crif3\RULE60\09863475.txt
Output Set: N:\CRF3\09252001\I863475.raw

164	65	70	75	80
166	Cys Ser Glu Met Val Pro Gly Thr Ala Asp Cys His Ile Thr Ala Asp			95
167		85	90	
169	Arg Lys Val Tyr Pro Gln Ala Asp Thr Val Ile Val His His Trp Asp			110
170		100	105	
172	Ile Met Ser Asn Pro Lys Ser Arg Leu Pro Pro Ser Pro Arg Pro Gln			125
173		115	120	
175	Gly Gln Arg Trp Ile Trp Phe Asn Leu Glu Pro Pro Asn Cys Gln			140
176		130	135	
178	His Leu Glu Ala Leu Asp Arg Tyr Phe Asn Leu Thr Met Ser Tyr Arg			160
179		145	150	
181	Ser Asp Ser Asp Ile Phe Thr Pro Tyr Gly Trp Leu Glu Pro Trp Ser			175
182		165	170	
184	Gly Gln Pro Ala His Pro Pro Leu Asn Leu Ser Ala Lys Thr Glu Leu			190
185		180	185	
187	Val Ala Trp Ala Val Ser Asn Trp Lys Pro Asp Ser Ala Arg Val Arg			205
188		195	200	
190	Tyr Tyr Gln Ser Leu Gln Ala His Leu Lys Val Asp Val Tyr Gly Arg			220
191		210	215	
193	Ser His Lys Pro Leu Pro Lys Gly Thr Met Met Glu Thr Leu Ser Arg			240
194		225	230	
196	Tyr Lys Phe Tyr Leu Ala Phe Glu Asn Ser Leu His Pro Asp Tyr Ile			255
197		245	250	
199	Thr Glu Lys Leu Trp Arg Asn Ala Leu Glu Ala Trp Ala Val Pro Val			270
200		260	265	
202	Val Leu Gly Pro Ser Arg Ser Asn Tyr Glu Arg Phe Leu Pro Pro Asp			285
203		275	280	
205	Ala Phe Ile His Val Asp Asp Phe Gln Ser Pro Lys Asp Leu Ala Arg			300
206		290	295	
208	Tyr Leu Gln Glu Leu Asp Lys Asp His Ala Arg Tyr Leu Ser Tyr Phe			320
209		305	310	
211	Arg Trp Arg Glu Thr Leu Arg Pro Arg Ser Phe Ser Trp Ala Leu Asp			335
212		325	330	
214	Phe Cys Lys Ala Cys Trp Lys Leu Gln Glu Ser Arg Tyr Gln Thr			350
215		340	345	
217	Val Arg Ser Ile Ala Ala Trp Phe Thr			360
218		355		

220 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1500 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

234	CCTTCCCTTG TAGACTCTTC TTGGAATGAG AAGTACCGAT TCTGCTGAAG ACCTCGCGCT	60
236	CTCAGGCTCT GGGAGTTGGA ACCCTGTACC TTCCTTTCCT CTGCTGAGCC CTGCCTCCTT	120
238	AGGCAGGCCA GAGCTCGACA GAACTCGGTT GCTTTGCTGT TTGCTTTGGA GGAACACAG	180

RAW SEQUENCE LISTING

DATE: 09/25/2001

PATENT APPLICATION: US/09/863,475

TIME: 10:33:32

Input Set : N:\Crif3\RULE60\09863475.txt

Output Set: N:\CRF3\09252001\I863475.raw

```

242 CTGACGATGA GGCTGACTTT GAACTCAAGA GATCTGCTTA CCCCAGTCTC CTGGAATTAA      240
244 AGGCCTGTAC TACATTTGCC TGGACCTAAG ATTTTCATGA TCACTATGCT TCAAGATCTC      300
246 CATGTCAACA AGATCTCCAT GTCAAGATCC AAGTCAGAAA CAAGTCTTCC ATCCTCAAGA      360
248 TCTGGATCAC AGGAGAAAAT AATGAATGTC AAGGGAAAAG TAATCCTGTT GATGCTGATT      420
250 GTCTCAACCG TGGTTGTCGT GTTTTGGGAA TATGTCAACA GAATTCCAGA GGTGGTGAG      480
252 AACAGATGGC AGAAGGACTG GTGGTTCCCA AGCTGGTTTA AAAATGGGAC CCACAGTTAT      540
254 CAAGAAGACA ACGTAGAAGG ACGGAGAGAA AAGGGTAGAA ATGGAGATCG CATTGAAGAG      600
256 CCTCAGCTAT GGGACTGGTT CAATCCAAAG AACC GCCCGG ATGTTTTGAC AGTGACCCCG      660
258 TGGGAAGGCG CGATTGTGTG GGAAGGCACT TATGACACAG CTCTGCTGGA AAAGTACTAC      720
260 GCCACACAGA AACTCACTGT GGGGCTGACA GTGTTTGCTG TGGGAAAAGTA CATTGAGCAT      780
262 TACTTAGAAG ACTTTCTGGA GTCTGCTGAC ATGTACTTCA TGGTTGGCCA TCGGGTCATA      840
264 TTTTACGTCA TGATAGACGA CACCTCCCGG ATGCCTGTCTG TGCACCTGAA CCTCTACAT      900
266 TCCTTACAAG TCTTTGAGAT CAGGTCTGAG AAGAGGTGGC AGGATATCAG CATGATGCGC      960
268 ATGAAGACCA TTGGGGAGCA CATCCTGGCC CACATCCAGC ACGAGGTCGA CTTCTCTTTC     1020
270 TGCATGGACG TGGATCAAGT CTTTCAAGAC AACTTCGGGG TGGAAACTCT GGGCCAGCTG     1080
272 GTAGCACAGC TCCAGGCCTG GTGGTACAAG GCCAGTCCCG AGAAGTTCAC CTATGAGAGG     1140
274 CGGGAAGTGT CGGCCGCGTA CATTCATTC GGAGAGGGGG ATTTTACTA CCACGCGGCC     1200
276 ATTTTGGAG GAACGCTAC TCACATTCTC AACCTCACCA GGGAGTGCTT TAAGGGGATC     1260
278 CTCCAGGACA AGAAACATGA CATAGAAGCC CAGTGGCATG ATGAGAGCCA CCTCAACAAA     1320
280 TACTTCCTTT TCAACAAACC CACTAAAATC CTATCTCCAG AGTATTGCTG GGACTATCAG     1380
282 ATAGGCCTGC CTTGAGATAT TAAAAGTGTC AAGGTAGCTT GGCAGACAAA AGAGTATAAT     1440
284 TTGGTTAGAA ATAATGTCTG ACTTCAAATT GTGATGGAAA CTTGACACTA TTTCTAACCA     1500

```

287 (2) INFORMATION FOR SEQ ID NO: 4:

289 (i) SEQUENCE CHARACTERISTICS:

290 (A) LENGTH: 394 amino acids

291 (B) TYPE: amino acid

292 (D) TOPOLOGY: unknown

294 (ii) MOLECULE TYPE: protein

298 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

300 Met Ile Thr Met Leu Gln Asp Leu His Val Asn Lys Ile Ser Met Ser
301 1 5 10 15
303 Arg Ser Lys Ser Glu Thr Ser Leu Pro Ser Ser Arg Ser Gly Ser Gln
304 20 25 30
306 Glu Lys Ile Met Asn Val Lys Gly Lys Val Ile Leu Leu Met Leu Ile
307 35 40 45
309 Val Ser Thr Val Val Val Val Phe Trp Glu Tyr Val Asn Arg Ile Pro
310 50 55 60
312 Glu Val Gly Glu Asn Arg Trp Gln Lys Asp Trp Trp Phe Pro Ser Trp
313 65 70 75 80
315 Phe Lys Asn Gly Thr His Ser Tyr Gln Glu Asp Asn Val Glu Gly Arg
316 85 90 95
318 Arg Glu Lys Gly Arg Asn Gly Asp Arg Ile Glu Glu Pro Gln Leu Trp
319 100 105 110
321 Asp Trp Phe Asn Pro Lys Asn Arg Pro Asp Val Leu Thr Val Thr Pro
322 115 120 125
324 Trp Lys Ala Pro Ile Val Trp Glu Gly Thr Tyr Asp Thr Ala Leu Leu
325 130 135 140
327 Glu Lys Tyr Tyr Ala Thr Gln Lys Leu Thr Val Gly Leu Thr Val Phe
328 145 150 155 160

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/863,475

DATE: 09/25/2001
TIME: 10:33:32

Input Set : N:\Crif3\RULE60\09863475.txt
Output Set: N:\CRF3\09252001\I863475.raw

330 Ala Val Gly Lys Tyr Ile Glu His Tyr Leu Glu Asp Phe Leu Glu Ser
331 165 170 175
333 Ala Asp Met Tyr Phe Met Val Gly His Arg Val Ile Phe Tyr Val Met
334 180 185 190
336 Ile Asp Asp Thr Ser Arg Met Pro Val Val His Leu Asn Pro Leu His
337 195 200 205
339 Ser Leu Gln Val Phe Glu Ile Arg Ser Glu Lys Arg Trp Gln Asp Ile
340 210 215 220
342 Ser Met Met Arg Met Lys Thr Ile Gly Glu His Ile Leu Ala His Ile
343 225 230 235 240
345 Gln His Glu Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe
346 245 250 255
348 Gln Asp Asn Phe Gly Val Glu Thr Leu Gly Gln Leu Val Ala Gln Leu
349 260 265 270
351 Gln Ala Trp Trp Tyr Lys Ala Ser Pro Glu Lys Phe Thr Tyr Glu Arg
352 275 280 285
354 Arg Glu Leu Ser Ala Ala Tyr Ile Pro Phe Gly Glu Gly Asp Phe Tyr
355 290 295 300
357 Tyr His Ala Ala Ile Phe Gly Gly Thr Pro Thr His Ile Leu Asn Leu
358 305 310 315 320
360 Thr Arg Glu Cys Phe Lys Gly Ile Leu Gln Asp Lys Lys His Asp Ile
361 325 330 335
363 Glu Ala Gln Trp His Asp Glu Ser His Leu Asn Lys Tyr Phe Leu Phe
364 340 345 350
366 Asn Lys Pro Thr Lys Ile Leu Ser Pro Glu Tyr Cys Trp Asp Tyr Gln
367 355 360 365
369 Ile Gly Leu Pro Ser Asp Ile Lys Ser Val Lys Val Ala Trp Gln Thr
370 370 375 380
372 Lys Glu Tyr Asn Leu Val Arg Asn Asn Val
373 385 390

375 (2) INFORMATION FOR SEQ ID NO: 5:

377 (i) SEQUENCE CHARACTERISTICS:

- 378 (A) LENGTH: 8174 base pairs
379 (B) TYPE: nucleic acid
380 (C) STRANDEDNESS: unknown
381 (D) TOPOLOGY: unknown

383 (ii) MOLECULE TYPE: DNA (genomic)

385 (iv) ANTI-SENSE: NO

389 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

391	GAATTCCATC	GTGGCAAGGG	CAGCCTGAAT	GGATGATGTA	ACCTGGGGTC	CTTTCAATGG	60
393	AGGGCCAGAC	TCCTGGGTCT	AGGGGATGAG	GGAGGGGAGG	ATCGGGTTAG	CTGGGACCCA	120
395	GGTGAAAGGG	GCTGGGGGCT	CACATTCCTG	AGTCTCAGAG	AGAAGGATCT	GGGGTCTCAA	180
397	GCACCTGAGT	CGGAGGGAGG	AGGGGTGCTG	GGCTCCTGGA	AAAACCACCT	CTTGGACCAT	240
399	CTATGCAGAT	CACGCAGAAC	AAGAGAAATT	TCTGCGCCCC	ATCTGAATTT	CTAAGTTTGG	300
401	GGGGAGGGCG	TGATCTGACA	CTGAGGTTCC	TTGATCCTCA	GCAAGGCGGC	AATTGCTGTA	360
403	TGAAAGAAGC	GACCGCATCT	GAGACACAAG	TATCCTGCCT	TGGAAGCCTC	TCACCTGGCC	420
405	GTGGGCAAC	CTCAACCTCA	TCTGTCCCTG	CTCAGATGCT	CAGACCCTGG	ACATCCCAGC	480
407	CTCCTCTCC	CTGATGCAAT	CCTGGTGTTT	CTTTCACCA	AGAAGCCATC	CCAGGCCAG	540
409	GCAGGTGCTC	CTGAAATAAC	CTGGGGGGAG	GGGTGGCTGA	AAGTCCCTGA	CTGGAGTTGG	600

VERIFICATION SUMMARY

DATE: 09/25/2001

PATENT APPLICATION: US/09/863,475

TIME: 10:33:33

Input Set : N:\Crf3\RULE60\09863475.txt

Output Set: N:\CRF3\09252001\I863475.raw

L:31 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:32 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]